HIGH TEMPERATURE OXIDE DEPOSITION FOR EEPROM DEVICES

TECHNICAL FIELD

This invention relates, in general, to methods for fabricating semiconductor devices and, more particularly, to methods for fabricating two-bit EEPROM devices.

BACKGROUND ART

Non-volatile memory devices are currently in widespread use in electronic components that require the retention of information when electrical power is terminated. Non-volatile memory devices include read-only-memory (ROM), programmable-read-only memory (PROM), erasable-programmable-read-only memory (EPROM), and electrically-erasable-programmable-read-only-memory (EEPROM) devices. EEPROM devices differ from other non-volatile memory devices in that they can be electrically programmed and erased. Flash EEPROM devices are similar to EEPROM devices in that memory cells can be programmed and erased electrically. However, Flash EEPROM devices enable the erasing of all memory cells in the device using a single electrical current pulse.

Typically, an EEPROM device includes a floating-gate electrode upon which electrical charge is stored. The floating-gate electrode overlies a channel region residing between source and drain regions in a semiconductor substrate. The floating-gate electrode together with the source and drain regions forms an enhancement transistor. By storing electrical charge on the floating-gate electrode, the threshold voltage of the enhancement transistor is brought to a relatively high value. Correspondingly, when charge is removed from the floating-gate electrode, the threshold voltage of the enhancement transistor is brought to a relatively low value. The threshold level of the enhancement transistor determines the current flow through the transistor when the transistor is turned on by the application of appropriate voltages to the gate and drain. When the threshold voltage is high, no current will flow through the transistor, which is defined as a logic 0 state. Correspondingly, when the threshold voltage is low, current will flow through the transistor, which is defined as a logic 1 state.

In a flash EEPROM device, electrons are transferred to a floating-gate electrode through a dielectric layer overlying the channel region of the enhancement transistor. The electron transfer is initiated by either hot electron injection, or by Fowler-Nordheim tunneling. In either electron transfer mechanism, a voltage potential is applied to the floating-gate by an overlying control-gate electrode. The control-gate electrode is

capacitively coupled to the floating-gate electrode, such that a voltage applied on the control-gate electrode is coupled to the floating-gate electrode. The flash EEPROM device is programmed by applying a high positive voltage to the control-gate electrode, and a lower positive voltage to the drain region, which transfers electrons from the channel region to the floating-gate electrode. The flash EEPROM device is erased by grounding the control-gate electrode and applying a high positive voltage through either the source or drain region of the enhancement transistor. Under erase voltage conditions, electrons are removed from the floating-gate electrode and transferred into either the source or drain regions in the semiconductor substrate.

Product development efforts in EEPROM device technology have focused on increasing the programming speed, lowering programming and reading voltages, increasing data retention time, reducing cell erasure times and reducing cell dimensions. Many of the foregoing research goals can be addressed through development of materials and processes for the fabrication of the floating-gate electrode. Recently, development efforts have focused on dielectric materials for fabrication of the floating-gate electrode. Silicon nitride in combination with silicon dioxide is known to provide satisfactory dielectric separation between the control-gate electrode and the channel region of the enhancement transistor, while possessing electrical characteristics sufficient to store electrical charge.

One important dielectric material for the fabrication of the floating-gate electrode is an oxide-nitride-oxide (ONO) structure. During programming, electrical charge is transferred from the substrate to the silicon nitride layer in the ONO structure. Voltages are applied to the gate and drain creating vertical and lateral electric fields, which accelerate the electrons along the length of the channel. As the electrons move along the channel, some of them gain sufficient energy to jump over the potential barrier of the bottom silicon dioxide layer and become trapped in the silicon nitride layer. Electrons are trapped near the drain region because the electric fields are the strongest near the drain. Reversing the potentials applied to the source and drain will cause electrons to travel along the channel in the opposite direction and be injected into the silicon nitride layer near the source region. Because silicon nitride is not electrically conductive, the charge introduced into the silicon nitride layer tends

to remain localized. Accordingly, depending upon the application of voltage potentials, electrical charge can be stored in discrete regions within a single continuous silicon nitride layer.

Various memory devices incorporating the features described above are known in the art. For example, WO 82/40162 describes the features recited in the preamble of the claims. The German patent DE 3032364 also describes a method for manufacturing a floating gate memory cell incorporating an oxide-nitride-oxide structure. US Patent 5104819 describes a method of forming an oxide-nitride-oxide composite dielectric structure between a floating polysilicon electrode and a control electrode in an EEPROM type device which increases the capacitive coupling between the control gate and the floating gate while providing greater charge retention than previous methods.

The article "Thin stacked oxide/nitride/oxide dielectrics formation by *in situ* multiple reactive rapid thermal processing" (Applied Physics Letters Vol. 55, No 22. p2313-2315) characterises an oxide-nitride-oxide dielectric formed in situ using a chemical vapour deposition technique.

Non-volatile memory designers have taken advantage of the localized nature of electron storage within a silicon nitride layer and have designed memory circuits that utilize two regions of stored charge within an ONO layer. This type of non-volatile memory device is known as a two-bit EEPROM. The two-bit EEPROM is capable of storing twice as much

information as a conventional EEPROM in a memory array of equal size. A left and right bit is stored in physically different areas of the silicon nitride layer, near left and right regions of each memory cell. Programming methods are then used that enable two-bits to be programmed and read simultaneously. The two-bits of the memory cell can be individually erased by applying suitable erase voltages to the gate and to either the source or drain regions.

While the recent advances in EEPROM technology have enabled memory designers to double the memory capacity of EEPROM arrays using two-bit data storage, numerous challenges exist in the fabrication of material layers within these devices. In particular, the ONO layer must be carefully fabricated to avoid the creation of interface states that could provide charge leakage paths within the ONO layer. Accordingly, advances in ONO fabrication technology are necessary to insure proper charge isolation in ONO structures used in two-bit EEPROM devices.

DISCLOSURE OF INVENTION

The present invention is for a process for fabricating an ONO floating-gate electrode in a two-bit EEPROM device. Fabrication of a two-bit EEPROM device using an ONO floating-gate electrode requires the formation of a high quality ONO structure. This is because proper functionality of the two-bit EEPROM device requires localized charge storage within the ONO structure. In particular, the top oxide layer must have a density sufficient to minimize the formation of charge traps. The formation of charge traps in the top oxide layer can lead to undesirable charge leakage within the top oxide layer and at the interface between the top oxide layer and the underlying silicon nitride layer. In a properly formed ONO structure, all electrical charge is stored in the silicon nitride layer. By fabricating a high quality top oxide layer, stored charge in the ONO structure remains localized within predetermined regions in the silicon nitride layer.

In one form, a process for fabricating an ONO floating-gate electrode includes providing a semiconductor substrate, and thermally growing a first silicon oxide layer on the semiconductor substrate. Then, a silicon nitride layer is formed to overlie the first silicon oxide layer. A second layer of silicon oxide is then formed to overlie the silicon nitride layer using a high temperature oxide (HTO) deposition process. The high temperature oxide deposition process is carried out at a temperature of about 700 to about 800° C.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 illustrates, in cross-section, a portion of a semiconductor substrate containing a floating-gate transistor that incorporates an ONO structure fabricated in accordance with the invention; and

FIGS. 2-4 illustrate, in cross-section, process steps for the fabrication of an ONO structure in accordance with the invention.

It should be appreciated that for simplicity and clarity of illustration, elements shown in the Figures have not necessarily been drawn to scale. For example, the dimensions of some of the elements are exaggerated relative to each other for clarity. Further, where considered appropriate, reference numerals have been repeated among the Figures to indicate corresponding elements.

MODES FOR CARRYING OUT THE INVENTION AND INDUSTRIAL APPLICABILITY

Shown in FIG. 1, in cross-section, is a floating-gate transistor 10 suitable for use in a two-bit EEPROM device. Floating-gate transistor 10 includes source/drain regions 12 and 14 located in a semiconductor substrate 16 and separated by a channel region 18. First and second bit line oxide regions 20 and 22 overlie source/drain regions 12 and 14, respectively. A control gate electrode 24 overlies channel region 18 and is separated therefrom by an ONO structure 26. Control gate electrode 24 and ONO structure 26 form a stacked-gate structure.

ONO structure 26 includes a first silicon oxide layer 28 overlying channel region 18. A silicon nitride layer 30 overlies first silicon oxide layer 28. A second silicon oxide layer (or top oxide layer) 32 overlies silicon nitride layer 30.

In the operation of floating-gate transistor 10, voltages are applied to control-gate electrode 24 and to source/drain regions 12 and 14 that cause electrical charge from source/drain regions 12 and 14 to propagate across channel region 18. Once the charge encounters a sufficiently strong vertical field, the charge is either injected or tunnels from channel region 18 into silicon nitride layer 30. For example, depending upon the particular voltage levels applied to control-gate electrode 24 and to source/drain regions 12 and 14, electrical charge 34 is transferred into silicon nitride layer 30 and is localized to regions in proximity to either source/drain region 12, or source/drain region 14.

Those skilled in the art will recognize that the proper functioning of a two-bit EEPROM device necessarily requires that electrical charge 34 remain isolated in the regions of silicon nitride layer 30 to which it is initially introduced. The proper maintenance of electrical charge 34 in localized regions of silicon nitride layer 30 is critical to the proper performance of a two-bit EEPROM device. In particular, the quality of ONO structure 26

must be such that charge leakage paths are minimized at the interface between the first and second silicon oxide layers 28 and 32, and silicon nitride layer. Additionally, second silicon oxide layer 32 must be of sufficient density, such that charge trapping sites are minimized within the silicon oxide material.

In accordance with the invention, charge leakage within ONO structure 26 is minimized by forming a high quality top oxide layer. The reduced charge leakage and improved floating-gate transistor performance obtained by the present invention can be better understood following a description of an ONO fabrication process carried out in accordance with the invention.

Referring to FIG. 2, first silicon dioxide layer 28 is formed to overlie the surface of semiconductor substrate 16. Preferably, semiconductor substrate 16 is a single crystal silicon substrate. Semiconductor substrate 16 has an upper surface 36 previously processed to remove debris and native oxides. Preferably, first silicon oxide layer 28 is formed by thermally oxidizing surface 36 at an elevated temperature in the presence of dry molecular oxygen. Preferably, the oxidation process is carried out at a temperature of about 900 to about 1100° C. The oxidation process forms a silicon oxide layer preferably having a thickness of about 5 to about 15 nanometers, and more preferably a thickness of about 10 nanometers. The oxidation process can be carried out in either a batch-type thermal oxidation furnace, or alternatively, in a single-wafer oxidation apparatus.

After forming first silicon oxide layer 28, silicon nitride layer 30 is formed to overlie first silicon oxide layer 28, as illustrated in FIG. 3. Preferably, silicon nitride layer 30 is formed by means of a rapid-thermal-chemical-vapor-deposition (RTCVD) process. The RTCVD process is carried out at a temperature of about 700 to about 800° C. The silicon nitride material is formed by reacting ammonia (NH_3) with either dichlorosilane ($SiCl_2H_2$) or silane (SiH_4). The processes is carried out for a period of time and at gas flow rates sufficient to form a silicon nitride layer preferably having a thickness of about 5 to about 15 nanometers, and more preferably about 100 angstroms. In one embodiment of the invention, ammonia is introduced in the RTCVD apparatus at a flow rate of about 1 standard liter per minute (slpm) and either dichlorosilane or silane is introduced at a flow rate of about 30 to about 50 standard-cubic-centimeters-per-minute (sccm). The RTCVD process is carried out in three steps including an initial temperature ramp, a deposition step, and cool-down step. Preferably, the total residence time of substrate 16 in the RTCVD apparatus is about 3 minutes. Preferably, the silicon nitride deposition step is completed in about 2 minutes.

Alternatively, silicon nitride layer 30 can be formed by means of a low-pressure-chemical-vapor-deposition (LPCVD) process. In the alternative method, the

silicon nitride can be formed in a batch deposition apparatus. The LPCVD process is preferably carried out at a pressure of about 200 to about 6666.13 kPa, at temperatures 700-800° C. using ammonia and either dichlorosilane or silane gas.

After forming silicon nitride layer 30, second silicon oxide layer 32 is formed to overlie silicon nitride layer 30, as illustrated in FIG. 4. In accordance with the invention, second silicon oxide layer 32 is formed by either an RTCVD or an LPCVD process. The second silicon oxide layer 32 is formed in an RTCVD apparatus using nitrous oxide (N_2O) and dichlorosilane. The RTCVD processes preferably carried out at a temperature of about 700 to about 800°C and for a total processing time of about 3 minutes. In similarity with the RTCVD process used to deposit silicon nitride layer 30, a three-step deposition sequence is used including a ramp up step, a deposition step, and a cool down step. The total processing time in the RTCVD apparatus is about 3 minutes. Preferably, during the deposition step about 1 to about 3 slpm, and more preferably, about 2 slpm of nitrous oxide is used. Additionally, about 25 to about 75 sccm, and more preferably, about 50 sccm of dichlorosilane is used. The RTCVD process forms a silicon oxide layer having a preferred thickness of about 5 to about 15 nanometers, and more preferably about 10 nanometers. In the LPCVD process, the second oxide is deposited on nitride in a batch type of furnace at temperatures of 700-800°C.

The foregoing RTCVD process for the formation of both the silicon nitride layer and the top oxide layer advantageously provides an in-situ process for the sequential formation of individual layers within the ONO structure. In particular, the formation of silicon nitride layer 30 and second silicon oxide layer 32 in a sequential, in-situ deposition sequence, minimizes contamination at the silicon nitride/silicon oxide interface. Additionally, uncontrolled native oxidation can be reduced to insure that a single, high-density silicon oxide layer is deposited over silicon nitride layer 30.

An important aspect of the invention includes the sequential formation of silicon nitride layer 30 and second silicon oxide layer 32 in the absence of exposure of silicon nitride layer 30 to ambient atmosphere. Following the deposition of silicon nitride layer 30 onto first silicon oxide layer 28, substrate 16 is transferred to an oxide deposition chamber under vacuum conditions without exposing the substrate to ambient atmosphere. Alternatively, a positive-pressure inert gas atmosphere can be introduced during wafer transfer. The in-situ nitride and oxide deposition response can be carried out in a batch-type LPCVD system.

Those skilled in the art will appreciate that various means exist for insuring that silicon nitride layer 30 is not exposed to ambient atmosphere prior to the deposition of second silicon oxide layer 32. For example, a cluster/tool apparatus can be used in which

semiconductor substrate 16 is transferred from a nitride deposition chamber to an oxide deposition chamber in a continuous vacuum environment, or in a positive pressure inert gas atmosphere. Alternatively, the nitride deposition and oxide deposition processes can be sequentially carried out in a single deposition chamber that is configured to receive various processing gases. Accordingly, all such variations and modifications are contemplated in the present invention.

Following the formation of ONO structure 26, the stacked-gate structure shown in FIG. 1 is completed by depositing a layer of gate forming material overlying second silicon oxide layer 32. A lithographic patterning and etching process is then carried out to define control-gate electrode 34 and an ONO floating-gate electrode. Those skilled in the art will recognize that various gate-forming materials can be used to fabricate control-gate electrode 24. For example, control-gate electrode 24 can be formed with polycrystalline silicon, amphorous silicon, a refractory metal silicide, and the like.

Thus, there has been disclosed in accordance with the invention a process for fabricating an ONO floating-gate electrode in a two-bit EEPROM device that fully provides the advantages set forth above. Although the invention has been described and illustrated with reference to specific illustrative embodiments thereof, it is not intended that the invention be limited to those illustrative embodiments. Those skilled in the art will recognize that variations and modifications can be made without departing from the spirit of the invention. For example, the thicknesses of the individual layers making up the ONO structure can be varied from that described herein. It is therefore intended to include within the invention all such variations and modifications that fall within the scope of the appended claims and equivalents thereof.

CLAIMS:

1. A process for fabricating an ONO floating-gate electrode in a two-bit EEPROM device comprising:

providing a single crystal silicon semiconductor substrate (16);

thermally oxidizing the silicon substrate (16) to form a first silicon dioxide layer (28) on the substrate (16);

forming a silicon nitride layer (30) on the first silicon dioxide layer (28); and

depositing a second layer of silicon dioxide (32) on the silicon nitride layer (30) using a high temperature oxide deposition process; characterised in that:

the silicon nitride layer (30) is formed using rapid thermal chemical vapor deposition using 0.5 to 2 standard-litres-per-minute (slpm) ammonia and 20 to 50 standard-cubic-centimetres-per-minute (sccm) of a second gas selected from a group consisting of silane and dichlorosilane; and

the high temperature oxide deposition process is carried out at a temperature of 700°C to 800°C so as to maintain at least two charge isolation regions in the silicon nitride layer (30).

2. A process as claimed in claim 1, wherein depositing a second silicon dioxide layer (32) comprises low pressure chemical vapor deposition using nitrous oxide and the second gas.

3. A process as claimed in claim 2, wherein depositing a second silicon dioxide layer (32) comprises low pressure chemical vapor deposition using 0.5 to 2 slpm nitrous oxide and 10 to 50 sccm of the second gas.

4. A process as claimed in claim 2 or claim 3, wherein depositing a second dioxide layer (32) comprises the low pressure chemical vapor deposition using nitrous oxide and dichlorosilane at a ratio from 10:1 to 200:1 by volume.

5. A process as claimed in claim 1, wherein depositing a second silicon dioxide layer (32) comprises rapid thermal chemical vapor deposition.

6. A process as claimed in claim 5, wherein depositing a second silicon dioxide layer (32) comprises rapid thermal chemical vapor deposition using 1 to 3 slpm nitrous oxide and 25 to 75 sccm dichlorosilane to form the second silicon dioxide layer (32) having a thickness of 50 to 150 Å.

7. A process as claimed in claim 5 or claim 6, wherein depositing a second dioxide layer (32) comprises the rapid thermal chemical vapor deposition using nitrous oxide and dichlorosilane at a ratio from 13:1 to 120:1 by volume.

8. A process as claimed in any one of claims 1 to 7, wherein the at least two charge isolation regions are maintained in the silicon nitride layer (30) by forming the silicon nitride layer (30) and the second silicon dioxide layer (32) sequentially in the absence of exposure to ambient atmosphere.